Philosophers' Imprint

VOLUME 14, NO. 35
DECEMBER, 2014

DEFERENCE DONE RIGHT

Richard Pettigrew
Michael G. Titelbaum

Department of Philosophy, University of Bristol, UK
Department of Philosophy, University of Wisconsin–Madison

1. Introduction

There are many kinds of *epistemic experts* to which we might wish to defer in setting our credences. These include: highly rational agents, objective chances, our own future credences, our own current credences, and evidential (or logical) probabilities. But how, precisely, ought we defer to these experts? Exactly what constraint does a deference requirement place on an agent's credences at a particular time?

In this paper we consider three possible answers, inspired by three different principles that have been proposed for deference to objective chances. We consider how these options fare when applied to the other kinds of epistemic experts mentioned above. Besides assuming a baseline probabilism about rational credences, we are particularly interested in the following two desiderata:

- A deference principle should be consistent with both the agent's and the experts' updating by Conditionalization.
- A deference principle should permit agents to have various kinds of doubts about what's rationally required.

Of the three deference principles we consider, we argue that two of the options face insuperable difficulties meeting these desiderata. The third, on the other hand, fares well — at least when it is applied in a particular way.

2. Deferring to experts

We begin by setting out the problem in general terms, not assuming that we are dealing with any particular sort of epistemic expert. Let $\mathcal{L}$ be a language. Let $C_1, \ldots, C_n$ be a finite set of distributions over $\mathcal{L}$, which we'll call the *candidate distributions*. That is, each C_i takes each sentence in $\mathcal{L}$ and assigns it a real number in $[0, 1]$. We will assume the candidate distributions are probability functions. Let $e_1, \ldots, e_n$ be among the atomic sentences of $\mathcal{L}$. We will take the sentence e_i to express the proposition that candidate C_i is the true expert (with the relevant sense of expertise to be filled in for each of our applications).

Call these the *expert hypotheses*.[1] We assume that the candidates take the $e_1, \ldots, e_n$ to be mutually exclusive and exhaustive.[2]

Given a body of evidence E, call a candidate C_i *immodest in the presence of* E just in case $C_i(e_i \mid E) = 1$ — that is, C_i becomes certain of its own expertise upon supposing E. Call a candidate *modest in the presence of* E if it isn't immodest. Given a particular feature candidates might have, call a candidate C_i *tolerant of that feature in the presence of* E just in case there is C_j possessing the feature to which C_i assigns positive credence of expertise conditional on E — that is, $C_i(e_j \mid E) > 0$. So, for example, C_i is tolerant of immodesty in the presence of E just in case there's a C_j immodest in the presence of E such that $C_i(e_j \mid E) > 0$. (Henceforth we will suppress 'in the presence of E' where context makes the relevant E clear.) Notice that all immodest candidates are tolerant of immodesty.

Now suppose we also have an agent. Just like the candidates, the agent has a distribution over $\mathcal{L}$, which we'll label cr. We assume cr is a probability function. We additionally assume that the agent, like the candidates, takes the $e_1, \ldots, e_n$ to be mutually exclusive and exhaustive. Finally, we let E be the agent's total evidence.

Suppose that when the agent sets her credences, she should try to defer to whichever candidate is the expert. The agent may be uncertain which candidate is the true expert, so we cannot direct her simply to defer to that one. Instead, we need a requirement incorporating the agent's opinions about the candidates' expertise. How might we formulate this requirement precisely as a constraint on her credence function cr? Here are three possibilities:

(PX) For all $x \in \mathcal{L}$ and all C_i,

$$cr(x \mid e_i) = C_i(x \mid E)$$

providing $cr(e_i), C_i(E) > 0$. If $C_i(E) = 0$, then $cr(e_i) = 0$.

This generalizes David Lewis' Principal Principle to experts other than chance functions [Lewis, 1980]. If the agent assigns positive credence to C_i being the true expert and C_i assigns positive probability to the agent's evidence E, then the agent ought to set her credence in x conditional on C_i being the true expert to whatever value C_i assigns to x once it has been brought up to speed with E. An agent should assign no probability to a distribution being the true expert if that distribution assigns no probability to the agent's evidence being true. In that case, the agent's credence in x conditional on C_i being the true expert is undefined and (PX) imposes no further constraints.

On certain views of evidence, this latter condition may sound implausible. Suppose I have evidence E but I am uncertain of this fact. Then surely I am not obliged to rule out a distribution as a candidate expert on the grounds that it assigns no probability to E. This is true, but this is not the account of evidence we assume here. Rather, we assume, with mainstream Bayesian epistemology, that it is a rational requirement that an agent assign credence 1 to her evidence. This is a consequence of the Bayesian updating norm of Conditionalization, for instance.

(NX) For all $x \in \mathcal{L}$ and all C_i,

$$cr(x \mid e_i) = C_i(x \mid Ee_i)$$

providing $cr(e_i), C_i(Ee_i) > 0$. If $C_i(Ee_i) = 0$, then $cr(e_i) = 0$.

This generalizes Ned Hall's and Michael Thau's New Principle [Thau, 1994], [Hall, 1994]. If the agent assigns positive credence to

1. For the sake of clarity and an issue that may come up later, if e_i were put into a natural-language sentence, it wouldn't say something like "Distribution C_i is an expert"; instead it would say something like "The distribution that assigns 0.5 to a_1, 0.7 to a_2, etc. is an expert." In other words, the propositions represented by the e_i refer to the candidates by description, not by name.
2. That is, $C_k(e_i e_j) = 0$ for all $i \neq j$. And $C_k(e_1 \vee \ldots \vee e_n) = 1$.

C_i being the true expert and C_i assigns positive probability to the conjunction of the agent's total evidence E and C_i being the true expert, then the agent ought to set her credence in x conditional on C_i being the true expert to whatever value C_i assigns to x once it has been brought up to speed with *E and the fact that C_i is the expert*. An agent should assign no probability to a distribution being the true expert if that distribution assigns no probability to the conjunction of her total evidence E and C_i being the true expert. In that case, the agent's credence in x conditional on C_i being the true expert is undefined and (NX) imposes no further constraints.

(IX) For all $x \in \mathcal{L}$,

$$cr(x) = \sum_{i=1}^{n} cr(e_i) \cdot C_i(x \mid E)$$

providing $C_i(E) > 0$ for all $i = 1, \ldots, n$. If $C_i(E) = 0$, then the corresponding term $cr(e_i) \cdot C_i(x \mid E)$ — which is anyway undefined — is omitted from the sum.

This generalizes Jenann Ismael's General Recipe [Ismael, 2008]. The agent ought to set her credence in x to her expectation of the expert's value for x once the expert has been brought up to speed with the agent's total evidence.

The 'X's in the names of these principles are meant to evoke variables; depending on the type of expert one considers, these principles might become variants of van Fraassen's Reflection Principle [van Fraassen, 1984], Elga's guru principle [Elga, 2007], Christensen's Rational Reflection principle [Christensen, 2010], etc. Initially, we submit, each principle seems plausible for all these applications. They might even seem to be equivalent. But with the possibility of modest distributions on the table, they are not:

Proposition 2.1

1. *(PX) entails (IX). But no other entailment relations hold between the three principles.*
2. *However, if all candidates are immodest in the presence of evidence E, the three principles make the same demands on an agent with total evidence E.*
3. *Just as (PX) entails (IX), (NX) entails a constraint on the unconditional credences assigned by cr in much the same way:*

$$cr(x) = \sum_{i=1}^{n} cr(e_i) \cdot C_i(x \mid Ee_i)$$

(All proofs are given in the appendix.)

One might wonder why, in each case, the candidate distributions to which we defer are brought up to speed with the agent's total evidence at the time of deference. If these distributions are really candidates for being experts for our agent, surely they already have at least as much evidence as she does? Not necessarily. Ned Hall draws an illuminating distinction between treating a distribution as an *analyst expert* and as a *database expert* [Hall, 2004]. We defer to database experts because of the evidence they have; we defer to analyst experts because of their talents at analyzing evidence they are given and assigning credences on the basis of that analysis. Certainly an analyst expert may be worthy of our deference even if our evidence exceeds hers. But it can also be worth deferring to a database expert who lacks some evidence we've got. I defer to a meteorologist on matters of tomorrow's weather at least in part because she is a database expert — her evidence about today's weather, for instance, is more extensive than mine. But that doesn't mean she has all of my evidence about today's weather; I know the weather at my precise location, whereas she is unlikely to. Nonetheless, her evidence is more extensive than mine, so I defer to her.

For both analyst and database experts, one should bring them up to speed on one's extra evidence before deferring to their opinions.[3] (PX), (NX), and (IX) are designed to do that. Moreover, formulating the principles this way lends them extra generality. If the candidates possess the agent's total evidence E already, conditionalizing on E does not alter their distributions. So even if one wants to restrict an agent's attention to candidates with at least as much evidence as she, there is no harm in E's place in these principles.

In the next two sections, we examine the formal features of the three deference principles. For readers who want to skip the technical presentation of our results and head directly to their interpretation in section 5, those results are:

- (PX), (IX), and (NX) all require that an agent be certain that the true expert assigns her evidence positive probability.
- (PX) requires, furthermore, that the agent be certain that the true expert is immodest in the presence of E.
- (PX) is preserved by Conditionalization.
- (IX) requires, furthermore, that the agent be certain that the true expert is intolerant of immodesty in other candidates.
- (IX) is not preserved by Conditionalization.
- (NX) requires nothing more of our agent's opinions about the true expert than that she be certain that it assigns positive probability to her evidence; (NX) makes demands on her other credences only once she has set her credences in the various expert hypotheses.
- (NX) may fail to be preserved by Conditionalization when two candidates converge on the same posterior distribution. Absent such convergence, Conditionalization preserves (NX).

3. Compare Elga's discussion of "guru" vs. "expert" principles [Elga, 2007].

3. Modesty and deference

(PX) imposes constraints on the credences an agent may assign to a proposition conditional on an expert hypothesis. But these constraints entail another constraint.

Proposition 3.1 *Suppose C_i is modest in the presence of the agent's total evidence E — that is, $C_i(e_i \mid E) < 1$. Then (PX) entails that $cr(e_i) = 0$.*

That is, (PX) demands that our agent be certain that the expert is immodest in the presence of the agent's total evidence; the agent must assign no credence whatsoever to modest candidates. One consequence is that if all candidates are modest, there are no probabilistic credences that satisfy (PX).

As we noted above, provided at least some candidates are modest in the presence of the agent's total evidence, (IX) is a weaker constraint than (PX). We might hope, then, that it will not require the same certainty in immodesty required by (PX). (IX) does not require that particular certainty, but it requires another.

Proposition 3.2 *Suppose C_i is tolerant of immodesty in other candidates — that is, there is $C_j \neq C_i$ such that $C_j(e_j \mid E) = 1$ and $C_i(e_j \mid E) > 0$. Then (IX) entails that $cr(e_i) = 0$.*

That is, (IX) requires the agent to be certain that the expert rules out all immodest candidates other than himself.

By contrast, (NX) does not require an agent to withhold credence from any particular candidate. Indeed, for any non-negative $\lambda_1, \ldots, \lambda_n$ that sum to 1, there is cr satisfying (NX) that assigns credence λ_i to expert hypothesis e_i — that is, $cr(e_i) = \lambda_i$. Define:

$$cr(x) = \sum_{i=1}^{n} \lambda_i C_i(x \mid E e_i)$$

Then $cr(e_i) = \lambda_i$ for all i, and cr satisfies (NX).[4]

In the second half of the paper, we will argue that these features rule out (PX) and (IX) as deference principles for a number of interpretations of the candidate distributions. Before we do that, we consider the diachronic constraints imposed by our principles.

4. Updating and deference

Suppose that, at some initial time t, our agent's credences are given by cr. At that time, the distributions $C_1, \ldots, C_n$ are the candidate experts; and the proposition e_i expresses the expert hypothesis that *C_i is the expert distribution at t*. Now suppose that our agent gains some evidence E between t and a later time t'. Her credences at t' are given by cr'. At that later time, the distributions $D_1, \ldots, D_m$ are the candidate experts; and the proposition f_i expresses the expert hypothesis that *D_i is the expert distribution at t'*. One might wonder under what circumstances the following four conditions can simultaneously be met:

(i) The distributions $D_1, \ldots, D_m$ are obtained by conditionalizing the distributions $C_1, \ldots, C_n$ on E.[5]
(ii) cr satisfies one of our three deference principles with respect to the distributions $C_1, \ldots, C_n$.
(iii) cr' is obtained from cr by conditionalizing on E;
(iv) cr' satisfies the same deference principle, but with respect to the distributions $D_1, \ldots, D_m$.

We will now take up each of our deference principles in turn, and consider for that principle under what circumstances all four of these conditions can be met.

4.1 *(PX) and Conditionalization*

We'll start by taking (PX) as our deference principle.

Proposition 4.1 *Suppose cr satisfies (PX) at t, cr' is obtained from cr by updating on E, and the distributions $D_1, \ldots, D_m$ are obtained from the $C_1, \ldots, C_n$ by conditionalizing on E. Then cr' satisfies (PX) at t'.*

In other words, conditions (i), (ii), and (iii) above together entail condition (iv). When this happens, we say that Conditionalization preserves (PX): Conditionalization takes (PX) distributions to (PX) distributions.

4.2 *(IX) and Conditionalization*

Things do not work out so happily for (IX) and (NX). We consider (IX) first. There are situations in which cr satisfies (IX), cr' is obtained from cr by Conditionalization, but cr' does not satisfy (IX). This is because, in general, linear averaging doesn't commute with conditionalizing.[6] This is illustrated by the following example:

Example 1 *Suppose the atomic sentences of $\mathcal{L}$ are e_1, e_2, a_1, a_2.*

- *e_1 says that the candidate distribution C_1 is the true expert;*
- *e_2 says that the candidate distribution C_2 is the true expert;*
- *there is no constraint on the interpretation of a_1 or a_2 — they might be sentences about the weather, or about the outcome of a basketball game.*

The following table gives:

4. After all, for all e_k such that $cr(e_k) > 0$:

$$cr(x \mid e_k) = \frac{cr(xe_k)}{cr(e_k)} = \frac{\sum_{i=1}^{n} \lambda_i C_i(xe_k \mid Ee_i)}{\sum_{i=1}^{n} \lambda_i C_i(e_k \mid Ee_i)} = C_i(x \mid Ee_k)$$

as required.

5. For some interpretations of the candidate functions it will be implausible to suppose that the candidates and the agent update on the same proposition between two times. Suppose we are considering chance as an expert. Chances evolve by conditionalizing on whatever actually transpires between two times, which may be distinct from what the agent learns between those times. Nonetheless, the sort of situation we consider — in which the candidate and the agent learn exactly the same proposition between t and t' — *could* arise for any sort of expert. So it is legitimate to ask how our putative deference principles would treat such a situation and to judge them on the answer.

6. Cf. [Jehle and Fitelson, 2009] and [Lehrer and Wagner, 1981].

- *A prior distribution cr, the agent's distribution at t;*
- *Two distributions C_1 and C_2 that we will take to be the candidates at t. cr satisfies (IX) with respect to C_1 and C_2.*
- *The posterior distribution cr′, which is obtained from cr by conditionalizing on evidence a_2.*
- *Two distributions D_1 and D_2 that we will take to be the candidates at t′. They are obtained from C_1 and C_2 by conditionalizing on a_2.*

Each row of the table corresponds to a way that the world could be: e.g. $e_2\overline{a_1}a_2$ is the world at which e_2 is true, a_1 false, and a_2 true.

	cr	C_1	C_2	cr'	D_1	D_2
$e_1a_1a_2$	3/32	0	3/16	3/16	0	1/4
$e_1a_1\overline{a_2}$	3/32	3/16	0	0	0	0
$e_1\overline{a_1}a_2$	3/32	3/16	0	3/16	3/4	0
$e_1\overline{a_1}\overline{a_2}$	7/32	6/16	1/16	0	0	0
$e_2a_1a_2$	9/32	0	9/16	9/16	0	3/4
$e_2a_1\overline{a_2}$	1/32	1/16	0	0	0	0
$e_2\overline{a_1}a_2$	1/32	1/16	0	1/16	1/4	0
$e_2\overline{a_1}\overline{a_2}$	5/32	2/16	3/16	0	0	0

Let:

- *f_1 be the proposition that D_1 is the true expert at t′;*
- *f_2 be the proposition that D_2 is the true expert at t′.*

At t′, the agent is certain that $e_1 \equiv f_1$ and $e_2 \equiv f_2$. So we can use cr′ values in the e_i to determine cr′ values in the f_i. But then cr′ does not satisfy (IX) with respect to the candidates D_1 and D_2. After all:

$$cr'(a_1) = \frac{3}{4} \neq \frac{5}{8} = cr'(f_1)D_1(a_1) + cr'(f_2)D_2(a_1)$$

Thus, satisfaction of (IX) is not preserved by Conditionalization.

4.3 (NX) and Conditionalization

We have seen that Conditionalization preserves (PX) but not (IX). How does (NX) fare? The answer, it turns out, depends on the relationship between the candidates at t and the candidates at t'. Recall that $C_1, \ldots, C_n$ are the candidates at t, $D_1, \ldots, D_m$ are the candidates at t', and E is the evidence on which the candidates and the agent condition between t and t'. If we have $i \neq j$ such that $C_i(- \mid E) = C_j(- \mid E) = D_k(-)$, we will say that *$C_i$ and C_j converge to D_k upon receipt of E*.

It turns out that if no two candidates at the earlier time converge to a single candidate upon receipt of the evidence obtained by the later time, then (NX) is preserved by Conditionalization.

Proposition 4.2 *Suppose that, between t and t′, the candidates update by conditionalizing on E and no two candidates converge upon receipt of E. Then, if cr satisfies (NX), and cr′ is obtained from cr by conditionalizing on E, then cr′ satisfies (NX).*

However, this is not guaranteed if some of the candidates at t *do* converge upon receipt of the evidence obtained by t', as illustrated by the following example:[7]

Example 2 *Suppose C_1 and C_2 are the candidates at t and D is the candidate on which they converge at t′. As above, a_2 is both the evidence our agent learns and the proposition on which the earlier candidates update to obtain the later candidate. That is, $D(-) = C_1(- \mid a_2) = C_2(- \mid a_2)$. Also, as above, e_i says that C_i is the true expert at t, while f says that D is the true expert at t′. Thus, $f \equiv e_1 \vee e_2$.*

7. This phenomenon was brought to our attention by Grant Reaber.

	cr	C_1	C_2	cr'	D
$e_1a_1a_2$	2/17	1/9	1/6	26/137	2/9
$e_1a_1\overline{a_2}$	9/68	1/8	1/16	0	0
$e_1\overline{a_1}a_2$	2/17	1/9	1/6	26/137	2/9
$e_1\overline{a_1a_2}$	9/68	1/8	1/16	0	0
$e_2a_1a_2$	1/13	1/18	1/12	17/137	1/9
$e_2a_1\overline{a_2}$	3/52	1/8	1/16	0	0
$e_2\overline{a_1}a_2$	4/13	2/9	1/3	68/137	4/9
$e_2\overline{a_1a_2}$	3/52	1/8	1/16	0	0

Then cr satisfies (NX) with respect to C_1 and C_2. And cr' is obtained from cr by conditionalizing on a_2. But cr' does not satisfy (NX) with respect to D. After all, $cr'(f) = D(f) = D(e_1 \vee e_2) = 1$, yet

$$cr'(a_1 \mid f) = \frac{43}{137} \neq \frac{1}{3} = C_1(a_1 \mid a_2 f) = D(a_1 \mid f)$$

To give an indication of why such examples exist, consider why (PX) is preserved by conditionalization. First, note that $f \equiv e_1 \vee e_2$. Thus, we have

$$cr'(a_1 \mid f) = cr(a_1 \mid a_2 f) = cr(a_1 \mid a_2e_1 \vee a_2e_2)$$

Now, if cr obeys (PX), we have

$$cr(a_1 \mid a_2e_1) = C_1(a_1 \mid a_2) = D(a_1) = C_2(a_1 \mid a_2) = cr(a_1 \mid a_2e_2)$$

since C_1 and C_2 converge to D upon receipt of a_2. In general, if $c(X \mid A) = c(X \mid B)$ and A and B are mutually exclusive, then $c(X \mid A \vee B) = c(X \mid A) = c(X \mid B)$. So

$$cr'(a_1 \mid f) = cr(a_1 \mid a_2e_1 \vee a_2e_2) = cr(a_1 \mid a_2e_1) = cr(a_1 \mid a_2e_2) = D(a_1)$$

Thus cr' satisfies (PX) with respect to D. Why doesn't analogous reasoning establish that (NX) is preserved by conditionalization? The problem is that, in general, we don't have

$$cr(a_1 \mid a_2e_1) = cr(a_1 \mid a_2e_2)$$

If cr satisfies (PX), then $cr(a_1 \mid a_2e_i) = C_i(a_1 \mid a_2)$. If cr satisfies (NX), then $cr(a_1 \mid a_2e_i) = C_i(a_1 \mid a_2e_i)$. And, while we always have $C_1(a_1 \mid a_2) = C_2(a_1 \mid a_2)$ because C_1 and C_2 converge upon receipt of a_2, such convergence does not guarantee that $C_1(a_1 \mid a_2e_1) = C_2(a_1 \mid a_2e_2)$.

This completes our investigation of the formal consequences of (PX), (IX), and (NX). To repeat our earlier summary:

- (PX), (IX), and (NX) all require that an agent be certain that the true expert assigns her evidence positive probability.
- (PX) requires, furthermore, that the agent be certain that the true expert is immodest in the presence of E.
- (PX) is preserved by Conditionalization.
- (IX) requires, furthermore, that the agent be certain that the true expert is intolerant of immodesty in other candidates.
- (IX) is not preserved by Conditionalization.
- (NX) requires nothing more of our agent's opinions about the true expert than that she be certain that it assigns positive probability to her evidence; (NX) makes demands on her other credences only once she has set her credences in the various expert hypotheses.
- (NX) may fail to be preserved by Conditionalization when two candidates converge on the same posterior distribution. Absent such convergence, Conditionalization preserves (NX).

In the following sections, we'll investigate the implications of these formal facts for deference to different sorts of experts.

5. Deferring to outside experts

We will consider three sorts of expert to which an agent might defer: herself (either currently or in the future), other agents, and ideal agents.[8] We begin, in this section, by considering other agents.

5.1 Outside experts and modesty

You are attending a conference on climate science. In the room are some of the world's best climatologists. All have the same evidence, and they all have all the evidence you have. But some, you believe, are better than others at assessing that evidence. Indeed, you believe that one is better than all the others, and that's the one you want to set your credences by as far as propositions concerning the Earth's climate go. Unfortunately, you don't know which one it is.[9] Let $\mathcal{L}$ be a language. It includes sentences about Earth's climate. Let $C_1, \ldots, C_n$ be the climatologists' distributions over the sentences in $\mathcal{L}$. Let e_i be the sentence in $\mathcal{L}$ that says that C_i is the best climate scientist at the conference. As academics are wont to do, each of the candidates C_i assigns probabilities to each of the e_j. Your credences are given by cr, another distribution over the sentences of $\mathcal{L}$. You ought to do your best to defer to the best candidate. What constraint does that place on cr?[10]

Consider (PX). By Proposition 3.1, this demands that you assign no credence to any candidate C_i such that $C_i(e_i) < 1$. Thus, regardless of the candidate distributions you face in the room, you should be certain that whichever of the scientists is the best, she is herself certain that she is the best. That is, unless every scientist in the room is modest, in which case you must assign no credence to any of them being the best. Yet you are certain that one is the credences will not be best. So, in that situation, your additive.

(IX) improves the situation, but not by much. By Proposition 3.2, (IX) demands that you assign no credence to any candidate C_i for which there is $C_j \neq C_i$ such that $C_j(e_j) = 1$ and $C_i(e_j) > 0$. Thus, again regardless of the candidates in the room, you are certain that whichever scientist is best, she is certain that no colleague who is immodest is best.

Suppose Professor X thinks she may well be the best climatologist in the room; but she also entertains the possibility that in fact it is Professor Y; Professor Y, on the other hand, is absolutely certain that he is best. Does this rule out the possibility that Professor X is the best scientist in the room? It seems not.[11] But (IX) demands that you assign no credence to Professor X being the best climatologist in the room. Thus, (IX) demands of you that you assign no credence at all to a genuine possibility (thereby violating Regularity as well as common sense).

Of course, we are all familiar with cases in which plausible epistemic norms demand assigning no credence at all to a genuine possibility. The principle of indifference demands this when an agent considers an infinite, fair lottery. But this is not one of those situations. In those situations, the mathematical representation of probability is to blame. (The Archimedean property of the reals entails that there is just no "room" for a positive probability small enough for our purposes.)

8. For a related treatment of the case of deference to chance, see Pettigrew [ta].
9. We realize it is somewhat artificial to imagine a situation with a unique best expert to whom you try to defer at the expense of all others. In most situations, the experts will be equally good, or each will have greater expertise in their particular specialty — glaciology, for instance, or the El Niño Southern Oscillation. We have nothing to say here about what to do in these situations: that is the preserve of judgement aggregation. Nonetheless, the sort of scenario we treat does arise: moreover, it provides a good parallel to other expert situations we will consider later, and is also a limiting case of various expertise situations one might find in real life.
10. Compare the discussion in Elga [2007]. Since the deference principles we're considering allow for the possibility that the true expert lacks some evidence you have, Elga would class them as "guru" principles rather than "expert" principles.
11. The literature on expert-opinion elicitation shows that it's a complex, contingent question how an expert's accuracy is related to ratings of his expertise (both by others and by himself). See, for instance, Koriat [2012] and Burgman et al. [2011].

But that is not the case here: it's not the mathematical representation that's to blame — it's the norm.

One might respond to our arguments against (PX) and (IX) by saying that these deference principles should not apply to the expert hypotheses themselves. Instead of being quantified over all sentences x in the language $\mathcal{L}$, (PX) and (IX) should be quantified only over those that do not contain the atomic e_is. Call this the *restricted-scope response*, since it is based on the claim that the expert hypotheses fall outside the scope of the deference principle in question.[12]

The first thing to say about the restricted-scope response is that, formally speaking, it works. If one restricts the scope of (PX) and (IX) in this way, they no longer have the consequences we identified. Indeed, restricted in this way, (PX) and (IX) are as permissive as (NX). For any non-negative reals $\lambda_1, \ldots, \lambda_n$ that sum to 1, there are credences that satisfy the restricted version of (PX) that assign credence λ_i to expert hypothesis e_i regardless of the modesty or otherwise of C_i. And similarly for the restricted version of (IX).

However, the second thing to say is that, philosophically speaking, this response fails. Of course, in the particular case we are considering — deference to the best climatologist at the conference — it might be that we should restrict our deference principle in the manner suggested: perhaps we have empirical evidence that climate scientists are poor judges of their own expertise and that of their peers. But there will be situations in which the outside expert to whom we defer is an expert not only about the subject matter in question, but also about their own expertise and that of their peers. And, whatever the correct account of deference is, it should be able to handle this case just as well as it handles the restricted case.

12. Although she is not responding to the problems posed here, Jenann Ismael suggests something like this response in the case of deference to objective chances [Ismael, 2008]. She suggests that chance functions do not assign chances to the chance hypotheses themselves; they assign them only to particular events. For a response to Ismael in that case, see Pettigrew [ta, §6]; and for Ismael's reply, see Ismael [ta].

Moreover, hypotheses about expertise are often correlated with non-expert hypotheses. You might plausibly believe that the true expert was the one trained at a certain school, or the one whose record of past predictions matches the facts in particular ways. That a particular candidate meets one of these descriptions would be a sentence in $\mathcal{L}$ distinct from the e_i. In extreme cases, each expert hypothesis will have a perfectly correlated non-e_i counterpart. For example, it might be that a particular magazine is about to publish the name of the true expert in the room, and both you and the candidates entertain opinions about whose name will be printed. When such strict biconditionals obtain, even restricted-scope deference principles will produce objectionable results. For instance, (PX) will require you to be certain that whoever's name is printed was certain that her name would appear.

5.2 *Outside experts and updating*

Consider again our conference of climate scientists. At the end of the conference, a new piece of evidence arrives: the Himalayan glaciers are melting faster than previously thought. All the climatologists hear the news; so do you. How should you and the climatologists update your credences in light of this new evidence? The orthodox Bayesian answer is that you should condition on it. However, while that updating policy will always be consistent with (PX) (Proposition 4.1), it isn't always consistent with (IX) or with (NX) (Examples 1 and 2) — that is, there are situations in which one cannot satisfy (IX) with respect to candidate experts at t, update by the same evidence as those experts between t and t', and also satisfy (IX) with respect to the candidate experts at t'; and similarly for (NX).[13]

One might reply: So much the worse for Conditionalization as a general updating rule! After all, we have retained it as the correct updating rule for the scientists. But, when one defers to an expert, one

13. Of course, it doesn't follow that (IX) and Conditionalization are, strictly speaking, inconsistent with one another in the sense that no agent could satisfy both. An agent could satisfy both by satisfying (IX) and never updating. But the terminology is nonetheless a helpful shorthand.

should not update by conditionalizing on one's new evidence; rather, one should update each candidate by conditionalizing on the new evidence, and then defer.

There are two problems with this. First, it is not obvious how one should weight the various candidates after updating. Should one weight them each by one's original credence in their expertise? That option is worrisome, as the new evidence might have some effect on your views of their expertise. In fact, before the new evidence arrived you might have suppositionally considered how your opinions of the candidates would change in light of reports from the Himalayas. Conditionalization directs you to be faithful to that reasoning once the new evidence arrives, but that's the option being ruled out. What other option is to be preferred?

A second problem: Conditionalization is not without justification. The diachronic Dutch book is one such argument. Another is a little-known point made by Peter M. Brown [Brown, 1976]. Suppose you know in advance that you will receive new evidence — perhaps you know that you will learn the rate at which the Himalayan glaciers are melting, though you don't know exactly which rate that will be. And suppose you know that, after being informed about the glaciers (but before gaining any extra evidence in addition to that), you will be required to make a decision between a range of possible actions — perhaps you will be asked to determine government climate policy. Suppose further that you know you will make that choice by maximizing expected utility with respect to the credences you assign at the time of the decision. Then your current expected utility of the action you will decide to perform is greatest if you believe that you will incorporate the glacier evidence by conditionalizing. Thus, if you adopt the alternative update-the-candidates-then-defer rule just proposed, you will expect the decisions you make on the basis of credences generated by that rule to be worse than the decisions you would have made had you conditioned. This is an unacceptable clash of epistemic and pragmatic norms. Another such clash is, of course, the vulnerability to diachronic Dutch books to which any intention to violate Conditionalization gives rise.

So we have good reason to preserve Conditionalization. Does this rule out (IX) and (NX) as principles of deference? We contend that it rules out (IX) but not (NX). In the case of (IX), there is nothing we can do to mitigate its conflict with the updating rule. But in the case of (NX), there is — that is the lesson of Proposition 4.2.

(NX) conflicts with Conditionalization in the following sort of situation: Before the evidence comes in, you defer to the climatologists' credences at that time in the manner proposed by (NX). Amongst these climatologists are two who disagree before they learn the evidence, but come to agree in all their opinions in light of the evidence. Perhaps these climatologists strongly disagreed about what the Himalayan report would say, but given that it says what it does, they agree upon the consequences. In our earlier terminology, the new evidence makes their credences converge. You then face a dilemma. You can either defer to the climatologists' updated credences in the manner proposed by (NX), or you can update your own credences by conditionalizing. But, as Example 2 shows, you can't always do both.

The problem arises because we demand that an agent defer at different times to different sets of candidates. Proposition 4.2 shows that the problem disappears if we demand that she defer to the same set of candidates at every time: if our agent defers to the same candidates at each time, then we can never have $f_i \equiv e_{i_1} \vee \ldots \vee e_{i_k}$ with $k > 1$, the situation that creates the incompatibility of (NX) and Conditionalization; rather, we will have $f_i \equiv e_i$, for all i. But what kind of distribution can there be such that it makes sense for the agent to constantly defer to that *same* distribution, even as her evidence changes over time? We submit that having identified a set of candidate experts, the agent should defer to those candidates' *initial* or *prior* or *ur*-credences — the credences the candidates had prior to acquiring any evidence whatsoever, the credences the candidates condition with their evidence at a given time to obtain their credences at that time.

The correct deference principle for outside experts is this: At any time, an agent ought to satisfy (NX) with respect to the possible prior

distributions of the candidates. Suppose C_1, ..., C_n are the possible priors. There will typically be more of these than there are candidate experts, since we may be uncertain not only about which of the candidates is the expert, but also for a given candidate what prior distribution she is using to set her current credences. Suppose further that e_i is the proposition that *C_i is the true expert's actual prior distribution*. Then our agent's credence in X conditional on C_i being the true expert's actual prior ought to be given by bringing C_i up to speed with the agent's evidence and the fact that C_i is the true expert prior. This resolves the conflict between Conditionalization and (NX) because the agent defers to the same set of candidates at all times. While the putative experts in front of her may gain evidence and shift their credences over time, the set of possible ur-credences lying behind those candidates remains constant.

One might wonder why deference to ur-credences at all times is not vulnerable to the problems identified in section 4.3. After all, ur-credence functions can surely converge. Indeed they can. And if our agent were to defer to the ur-credence functions of the candidates at one time and their updated converged credence functions at a later time, she would fall foul of Example 2. But that is not our proposal. Rather, we propose that she defer to the ur-credence functions of the candidate experts at *all* times, regardless of evidence. And that puts her in the situation covered by Proposition 4.2.

At first, it might seem that this move is appropriate only for deferring to analyst experts — experts to whom we defer because of their ability to analyse evidence and respond appropriately. After all, a candidate's prior embodies her analyst capabilities; it encodes her response to any evidential situation. But by definition it embodies none of her evidence. Thus, the proposal seems not to capture the deference we owe to a database expert, since it does not give us access to the evidence she has that we prize.

But that's not quite right. On the current proposal, even though we demand deference to the candidates' prior distributions, their current distributions do not leave the picture. The agent knows what the candidates' current distributions are. If we assume that the candidates' current distributions are generated by conditionalizing their priors on their total evidence, and if we assume those priors satisfy the Regularity principle, then a candidate's total evidence can be read off her current distribution by seeing which propositions she assigns a credence of 1. So when we take a candidate's prior and feed it all of the agent's current evidence (including the agent's current evidence about that candidate's credences), we are giving information to the prior about what the candidate currently knows. In addition, we are giving the prior all of the agent's information about whether what that candidate knows is to be trusted — whether that candidate has a database worth deferring to.

This point also addresses another worry about the approach. It might seem implausible that agents have opinions about the priors from which the candidates' current distributions have evolved. But if an agent knows both a candidate's current distribution and her current total evidence, the agent can reconstruct a great deal of information about that candidate's prior. This will be especially true when the agent's evidence yields other clues about how the candidate is likely to reason. In a room full of climatologists, it is highly likely that a candidate's opinions will have been shaped by certain general principles, that physically outlandish scenarios will be entertained only with extraordinary evidence, etc. A candidate's prior embodies her analyst function; the very fact that a climatologist was invited to the conference probably tells us a great deal about what kind of analyst she is.[14]

In sum: Neither (PX) nor (IX) provides the correct formulation of

14. It also helps here that for any given deferential situation, the agent need not reconstruct a candidate's *entire* prior — only the portions relevant to the propositions being deferred about will be of interest to the agent. With that said, an obvious extension of our work here would be to analyze deference principles for agents with "imprecise credences": levels of confidence modelled using a range of real numbers instead of a single value. That kind of analysis would open up more possibilities for modelling an agent's uncertainty about which particular prior belongs to a particular candidate.

the deference we owe to outside experts such as climatologists, stock market experts, or election pundits. Both principles require us to assign no credence at all to certain genuine possibilities; moreover, the latter principle is inconsistent with a well-motivated updating rule. (NX), on the other hand, imposes no constraints on the credences one might assign to expert hypotheses; and, applied to candidates' priors rather than their current distributions, it is always consistent with the updating rule.[15]

6. Rationally responding to the evidence

Some philosophers hold that, for *any* body of evidence, there is a unique credence function that is the rational response to that evidence. This is known as the Uniqueness Thesis [White, 2005; Feldman, 2007]. However, even if one does not accept the universal claim, all but the most stridently subjectivist Bayesians will agree that *there are* evidential situations admitting of only one rational response. Let us suppose that our agent is in such a situation, and let us suppose that she knows this. What she doesn't know is which credence function provides the rational response. But she has narrowed it down to a finite set $C_1, \ldots, C_n$. (How? Perhaps a Rationality Oracle has whispered the unique rational response in the agent's ear, but she didn't hear it properly. Or perhaps the agent knows that rationality will only ever require an agent to assign a credence precise to a particular number of decimal places, or expressible in a fraction with not-too-large whole numerator and denominator. . . .) As usual, let e_i be the sentence saying that C_i provides the unique rational response.

Clearly, we ought to defer to the unique rational response. So again the question arises how to formulate that requirement precisely.

15. One might wonder why we cannot remove the incompatibility of (IX) with Conditionalization in the same way that we removed the incompatibility of (NX) with Conditionalization, namely, by demanding that agents defer to an expert's *prior* or *initial* or *ur*-distribution. However, in Example 1, the two candidate experts at t — namely, C_1 and C_2 — do not converge to either of the candidate experts at t' — namely, D_1 or D_2. Thus, (IX) is incompatible with Conditionalization even when there is no convergence between t and t'.

6.1 *Evidential probabilities and modesty*

Consider (PX). This demands that our agent be certain that the correct rational response to her evidence is given by a candidate C_i such that $C_i(e_i) = 1$.[16] Now some philosophers hold that the truly rational response to a body of evidence will always be certain of its own rationality.[17] The idea is that whether or not a particular distribution provides the unique rational response to a given body of evidence is something knowable a priori. On an internalist account of justification, no information about the world should be required to know what the evidential probabilities are in the presence of a given body of evidence. So, if one is responding correctly to one's evidence, then one will know it — the true rational response to evidence is always immodest.

This apriorism is a controversial view. For (PX) to be plausible, not only must the view be true, but every evidential situation must require its agent to be *certain* that the view is true. By Proposition 3.1, an agent who satisfies (PX) rules out any candidate for the evidential probabilities that is not certain of its own expertise. So she eliminates the possibility that adopting the credences truly required by her evidence would leave her the slightest bit uncertain that she had done so. If we are concerned with epistemic modesty, that concern should apply equally to modesty about which epistemic theory is correct.

16. In this section we will mostly leave out the conditionalizing on E, as we assume that the *evidential probabilities* for an agent — the correct credences for her to assign *on her current evidence* — already incorporate that evidence.

17. See e.g. Titelbaum [ta]. A referee for this journal also helpfully pointed out that the models of epistemic probability in Williamson [2000] imply that facts about the evidential probabilities of particular hypotheses on particular bodies of evidence always have evidential probability 1. Now it may seem strange for this paper to draw out the consequences of a constraint one of its authors has argued against. Nevertheless, a number of philosophers (such as Elga [2013] and Christensen [2010]) have argued that evidential probabilities should sometimes be modest, and that seems to be the default intuitive position. So it's worth working out the consequences of such modesty for deference principles. Also, if no plausible deference principle could be made consistent with the modesty of evidential probabilities, some of us might construe that as a new argument against modesty.

Another way to put the objection is that (PX) is incompatible with the possibility of misleading higher-order evidence about what's rational. If such evidence is possible, we can have a situation in which the true evidential probabilities for an agent are captured by C_j but the agent's evidence leaves open the possibility that the evidential probabilties are C_i. Since the evidence leaves open the possibility of e_i, and C_j reflects what it's rational for the agent to believe on that evidence, we will have $C_j(e_i) > 0$. But then (PX) will prevent the agent from assigning any credence to e_j. In cases with misleading higher-order evidence, (PX) bars an agent from deferring to the credences that are rationally required by her evidence.[18]

(IX) does not prevent an agent from deferring to modest evidential probabilities, but does forbid deference if those probabilities are tolerant of immodesty. Thus a move from (PX) to (IX) in hopes of maintaining the possibility of rational modesty is ultimately self-defeating. The (IX) defender hopes to open up the possibility that some bodies of evidence could require an agent to be unsure whether she has responded rationally. But what of the agent's being unsure about whether such modesty is the rational response? Could an agent's evidence leave her rationally uncertain whether it's the kind of evidence that demands modesty or not? Certainly the modesty impulse moves us to admit the *possibility* of such cases. But on (IX), the only candidate responses to her evidence an agent can entertain that are modest about whether they are correct are at the same time absolutely certain that such modesty is required in response to the present evidence.

And again, (IX) is incompatible with certain kinds of misleading higher-order evidence. Suppose the evidential probabilities for an agent are captured by C_i, but her evidence leaves open the possibility that the evidential probabilities are some immodest C_j. In other words, $C_i(e_j) > 0$ and $C_j(e_j) = 1$. Then (IX) forbids the agent's assigning C_i positive credence. (IX) also fails as an account of deference to evidential probabilities.

As before, (NX) imposes neither of these implausible constraints on our credences in the various expert hypotheses; indeed, it imposes no constraints at all on such credences.

6.2 *Evidential probabilities and updating*

We know that (IX) and (NX) are not preserved by conditionalizing. On the accounts of deference that they provide, one might defer to the evidential probabilities at one time, then learn a new piece of evidence, condition on it and thereby fail to defer to the updated evidential probabilities at the later time. Nonetheless, we can make a similar move here to the one we made above to rescue (NX): When one updates on new evidence, one does not then defer to the updated evidential probabilities. Rather, at the earlier time and at the later time (and indeed at any time), one ought to defer to the prior evidential probabilities: that is, those distributions that are candidates for being the distribution we condition with a body of evidence to obtain the unique rational credences in the light of that evidence.

In this case, this move is extremely plausible. Evidential probabilities are the ultimate analyst experts. We defer to them not because of the evidence they possess — they always possess precisely what we possess — but because of their ability to assign appropriate credences in response to that evidence. Thus, it is natural to formulate the deference we owe to the evidential probabilities as follows: Our credence in a proposition X conditional on the prior evidential probabilities being given by C_i should be equal to C_i's probability for X conditional on our current evidence and the fact that C_i is the true evidential prior.

Once more, (NX) is the only deference principle that neither imposes implausibly strong synchronic constraints nor conflicts with plausible diachronic constraints.

18. Compare Elga's argument for his New Rational Reflection Principle over Christensen's Rational Reflection (which is essentially (PX) applied to evidential probabilities) [Elga, 2013].

7. Deferring to oneself

Finally, we turn to the deference one owes to oneself. This comes in two forms: one ought to defer to one's current credences, and one ought to defer to one's future credences. The former is the thought behind Christensen's principle of Self-Respect [Christensen, 2007, 322]; the latter is the thought behind van Fraassen's Reflection Principle [van Fraassen, 1984].

7.1 Deferring to one's current credences

Let us first consider deference to one's current credences. Some philosophers suggest that an agent ought to be certain, at a particular time, what her credences are at that time. Moran has suggested that the question 'Do I believe that p?' is transparent to the question 'Is p true? [Moran, 2001].[19] It's even harder to separate out the analogous questions concerning an agent's partial beliefs. Asking an agent to report her credence in p and asking her to assess what she takes her credence in p to be can each be accomplished with the query 'How confident are you that p?'. Any divergence between an agent's actual credence values and what she thinks those values might be could lead to strange courses of thought and behavior — for instance, if she sometimes accepts and rejects bets based on her actual credences but at other times acts on what she thinks her credences are.

But we can align an agent's first- and second-order opinions without demanding that her credences be *perfectly* transparent to her at all times. For example, one might think that a higher-order report is just a report of one's expectation of one's first-order credences. For that to line up with actual first-order credences would be for an agent to defer to her current credences in line with (IX) — where the candidates are possibilities one is entertaining for what one's own credence distribution might be, and $cr(e_i)$ gives one's credence that e_i represents one's actual opinions. The "true expert" is then just one's actual distribution.[20]

In this context, a modest candidate is a distribution that is uncertain what values it assigns; its values are not completely transparent to itself. A modest candidate intolerant of immodesty goes further, ruling out the very possibility that it's transparent. This candidate displays a form of negative introspection: its probabilities are not transparent and it is certain they are not. On the other hand, an immodest candidate is certain what values it assigns. Since it knows its own assignments, it is certain of its own immodesty (and therefore is intolerant of modesty). So an immodest candidate's probabilities display positive introspection: its probabilities are transparent and it is certain that they are so.

(IX) is not the only possibility for how an agent might defer to her own credences. Start with (PX). In this context, (PX) is the principle that Christensen calls Self-Respect; Skyrms calls it Miller's Principle, though that name is usually reserved for what is now called the Principal Principle [Skyrms, 1980, 112]. By Proposition 3.1, we see that it demands certainty in transparency — certainty that one has got one's own credences exactly right. This seems too strong. That is not to say that it is rationally prohibited to be certain of one's own transparency. Indeed, if an agent is certain of her own transparency, and she is right about this, then she can be certain she satisfies (PX).[21] But such certainty is not rationally compelled [Williamson, 2000].

So consider (IX). It says one's credences ought to match up with one's expectations of one's credences. Proposition 3.2 tells us that under (IX), an agent can assign positive credence only to a candidate that assigns no credence to other, immodest candidates. Thus all of the agent's credence will go either to immodest candidates (who assign no credence to other candidates of any type) or to modest candidates that

19. See also Evans [1982, p. 225].

20. As earlier, having finitely many candidates is plausible if we assume that real agents are capable of assigning credence values only to so many decimal points.

21. If $C_i(e_i) = 1$, then $C_i(X \mid e_i) = C_i(X)$. (Suppressing E again for obvious reasons.)

disregard immodest alternatives. In the current context, that means the agent will consider only distributions with positive introspection and distributions with negative introspection. She must not assign any positive credence to the possibility that her credences are not transparent and yet are unsure about their transparency. As before, this is too strong. Neither positive nor negative introspection is rationally compelled; so an agent can believe herself rational without being certain that her credences satisfy one of these two conditions.

Now consider the tension between (IX) or (NX) and Conditionalization in this context. Consider (IX). It turns out that, if (IX) and Conditionalization are both rational requirements, then an agent who understands what rationality requires and who believes herself to satisfy these rational requirements cannot in fact be rational. Let's see why that's so. If she is rational, she satisfies (IX) with respect to candidate distributions C_i at time t. Now suppose that, between t and the later time t', she receives some new evidence E. If she believes that she is rational and understands that Conditionalization is a rational requirement, she believes that her credences at t' are obtained from her credences at t by conditionalizing on E. Thus, her candidates at t' are D_1, ..., D_n, where $D_i(-) = C_i(- \mid E)$. But if she is rational she will have updated by Conditionalization, so her true credences will be given by $cr'(-) = cr(- \mid E)$. In that case she cannot satisfy (IX) at the later time t'. So she is not rational.

Similarly, suppose (NX) is the required deference rule, which the agent satisfied at t. If two of her C_is happen to converge to the same D_i under her new evidence, she cannot believe she is rational while obeying all that rationality requires.

Can we again rescue (NX) by suggesting that our deference be to priors rather than posterior distributions at a time? Yes. Christensen describes deferring to one's own credences as a mark of epistemic self-respect. This self-respect can be achieved as effectively by deferring to one's prior as it is by deferring to one's current credences, provided one's current credences are obtained by conditionalizing that prior. Having said that, it is surely possible for an agent to exhibit epistemic self-respect even if their current credence function has not been obtained from their ur-credence function by conditionalizing on their evidence. So something of the spirit of Christensen's principle is lost by this move.

7.2 *Deferring to one's future credences*

Finally, we consider how one might defer to one's future credences. In this case, (PX) is a version of van Fraassen's Reflection Principle, with the candidates being credence distributions the agent supposes she might assign at a particular future time. As usual, it demands that an agent be certain that her future credences are immodest. In this context, as in the previous section, immodesty amounts to transparency. Thus, (PX) demands that we assign no credence to the possibility that our future credences be less than transparent to us at that future time. But what grounds such confidence in our future transparency?

(IX), as before, demands less. We need not be certain of our future transparency. But we must be certain that either we will be transparent, or we will be less than transparent to ourselves and certain of this fact. Once more, this seems too strong a requirement.

More complicated is the question of how deference to our future credences interacts with Conditionalization. Suppose $C_1, \ldots, C_n$ are the distributions that our agent, at time t, thinks might give her credences at later time t''. Now suppose that between t and $t' < t''$, she learns E. She updates her credences cr at t to give her credences $cr'(-) = cr(- \mid E)$ at t'. But this does *not* mean she updates each $C_1, \ldots, C_n$ on E.

To see why, consider that there must be at least one C_i such that $C_i(E) < 1$. If each C_i were certain of E, then the agent would be certain at t that she was to learn E before t''. That would make the agent certain of E at t, meaning E could not count as new evidence for her at t'.

So at least some of the C_i do not have the agent learning E by t''. But since the agent has learned E at t', and $t' < t''$, the agent will rule out such C_i at t'. So the candidates for the agent at t' will be a proper subset of her candidates $C_1, \ldots, C_n$ from time t. Moreover, each of the remaining candidates will already incorporate E, so there is no need to condition them on E. The agent's candidates at t' really are just a proper subset of the candidates she considered at t.

The question now arises: Does cr' satisfy (IX) with respect to her candidates at t'? We need only tweak Example 1 to see that it need not.

Example 3 *Suppose that our agent learns $a_1a_2 \vee \overline{a_1a_2}$. cr' is obtained from cr by conditionalizing on this.*

	cr	C_1	C_2	cr'
$e_1a_1a_2$	3/32	0	3/16	1/8
$e_1a_1\overline{a_2}$	3/32	3/16	0	0
$e_1\overline{a_1}a_2$	3/32	3/16	0	0
$e_1\overline{a_1a_2}$	7/32	6/16	1/16	7/24
$e_2a_1a_2$	9/32	0	9/16	3/8
$e_2a_1\overline{a_2}$	1/32	1/16	0	0
$e_2\overline{a_1}a_2$	1/32	1/16	0	0
$e_2\overline{a_1a_2}$	5/32	2/16	3/16	5/24

The problem is that, by learning $a_1a_2 \vee \overline{a_1a_2}$, the agent has learned that C_2 must be her future credence. It is the only one of the two that is certain of $a_1a_2 \vee \overline{a_1a_2}$. Thus, by (IX), the agent ought to have credences that match the probabilities given by C_2. But, if she conditionalizes, they don't.

(NX), on the other hand, fares well. We need not even make the move to prior distributions here since, as pointed out above, none of the agent's candidates converge — they merely atrophy. Thus, (NX) is preserved by Conditionalization when we defer to our future credences.

8. Conclusion

We have examined three possible principles of deference to experts, based on three models found in the literature. While we make no claim that our list is exhaustive, of the options considered we find (NX) to be all-around best. On the one hand, it does not require implausibly strong certainties about the rational possibilities available to an agent. On the other hand, it is consistent with updating by Conditionalization, as long as deference is made to a candidate expert's priors instead of to her distribution posterior to the evidence.

We conclude by pointing out that (NX) may place constraints not only on an agent's credences, but also on the credences of experts to whom she should defer. Think back to the evidential probabilities application of (NX). Among the candidate prior evidential probability functions is the one true expert function. (The "true expert analyst," if you like.) For that function to be the true expert is just for it to dictate any agent's rational response to any possible course of evidence. So if the agent with credence function cr is rational, for any body of total evidence E she will have

$$cr(-) = C_j(- \mid E)$$

where C_j is the true expert function.

If (NX) is true, then for rational cr we have:

$$cr(x \mid e_i) = C_i(x \mid Ee_i)$$

Using the previous equation to expand the left-hand side yields

$$C_j(x \mid Ee_i) = C_i(x \mid Ee_i)$$

This must hold for any E; since that includes tautologous E, we can

generalize to

$$C_j(x \mid e_i) = C_i(x \mid e_i)$$

which is really just (NX) applied to an "agent" (the prior evidential probability function C_j) whose total evidence set is empty.

(NX) constrains not just the credences of the agent doing the deferring, but also the credences of the true expert among the candidates to whom she defers. Thus if the agent knows both (NX) and Conditionalization are correct, she can use (NX) to whittle down her list of candidate experts. But that shouldn't be surprising. (NX) is supposed to express a rational requirement; little wonder that it is satisfied by the perfectly rational credence distribution.[22]

Appendix A. Proofs

A.1 Proof of Proposition 2.1

1. Suppose $cr(x \mid e_i) = C_i(x \mid E)$. Then, since cr considers $e_1, \ldots, e_n$ to be exhaustive and mutually exclusive:

$$cr(x) = \sum_{i=1}^{n} cr(e_i) cr(x \mid e_i) = \sum_{i=1}^{n} cr(e_i) C_i(x \mid E)$$

2. Suppose $C_i(e_i \mid E) = 1$ for all $i = 1, \ldots, n$.

   - (PX) $\Leftrightarrow$ (NX). Since $C_i(e_i \mid E) = 1$, we have $C_i(x \mid Ee_i) = C_i(x \mid E)$. Thus, $cr(x \mid e_i) = C_i(x \mid E)$ iff $cr(x \mid e_i) = C_i(x \mid Ee_i)$.
   - (PX) $\Rightarrow$ (IX). See (1).
   - (IX) $\Rightarrow$ (PX). Suppose $cr(x) = \sum_{i=1}^{n} cr(e_i) \cdot C_i(x)$. Then

$$\begin{aligned} cr(x \mid e_k) &= \frac{cr(xe_k)}{cr(e_k)} \\ &= \frac{\sum_{i=1}^{n} cr(e_i) \cdot C_i(xe_k \mid E)}{\sum_{i=1}^{n} cr(e_i) \cdot C_i(e_k \mid E)} \\ &= \frac{C_k(xe_k \mid E)}{C_k(e_k \mid E)} \quad \text{since } C_i(e_k \mid E) = C_i(xe_k \mid E) = 0 \text{ if } i \neq k \\ &= C_k(x \mid Ee_k) = C_k(x \mid E) \end{aligned}$$

   This completes the proof. □

3. Similar to (1).

This completes our proof. □

A.2 Proof of Proposition 3.1

Suppose $cr(e_i) > 0$. Then $cr(e_i \mid e_i)$ is defined. Since cr is a probability function, $cr(e_i \mid e_i) = 1$. But, by (PX), $cr(e_i \mid e_i) = C_i(e_i \mid E) < 1$. This gives a contradiction and completes the proof. □

A.3 Proof of Proposition 3.2

Suppose (IX) and suppose that, for some i, there exists $j \neq i$ such that $C_j(e_j \mid E) = 1$ and $C_i(e_j \mid E) > 0$. Now suppose for *reductio* that $cr(e_i) > 0$.

By (IX),

$$cr(e_j) = \ldots + cr(e_i) \cdot C_i(e_j \mid E) + \ldots + cr(e_j) \cdot C_j(e_j \mid E) + \ldots$$

By assumption, $C_j(e_j \mid E) = 1$. So

$$cr(e_j) = \ldots + cr(e_i) \cdot C_i(e_j \mid E) + \ldots + cr(e_j) \cdot 1 + \ldots$$

22. We are very grateful to the following people for very helpful discussion of earlier versions of this paper: Adam Elga, Jenann Ismael, Grant Reaber, Joshua Schechter, and anonymous referees for this journal. Richard Pettigrew was supported by an ERC Starting Researcher Grant 'Epistemic Utility Theory: Foundations and Applications' during his work on this paper. Michael Titelbaum was supported by funds from the William F. Vilas Trust Estate.

Then, subtracting $cr(e_j)$ from both sides, we have

$$0 = \ldots + cr(e_i) \cdot C_i(e_j \mid E) + \ldots + 0 + \ldots$$

Since all the terms on the right-hand side of this equation must be non-negative (being products of probability values), this gives us

$$cr(e_i) \cdot C_i(e_j \mid E) = 0$$

But we supposed both these multiplicands were positive, so we have a contradiction.

□

A.4 Proof of Propositions 4.1 and 4.2.
Suppose

- cr is the agent's credence function at t. Her total evidence at t is E.
- cr' is the agent's credence function at t'. Her total evidence at t is E'.
- cr' is obtained from cr by conditionalizing on E. That is, $cr'(-) = cr(- \mid E')$.
- The candidates at t are $C_1, \ldots, C_n$.
- For each candidate C_i, $\mathcal{E}_i$ is the partition from which C_i will obtain evidence between t and t'.
- The candidates at t' are $D_1, \ldots, D_m$.

Thus, for each D_i, there are pairs C_j and $X \in \mathcal{E}_j$ such that $D_i(-) = C_j(- \mid X)$. So, if e_i says that C_i is the true expert at t, and f_i says that D_i is the true expert at t', then

$$f_i \equiv \bigvee_{e_j, X \in \mathcal{E}_j : D_i(-) = C_j(- \mid X)} e_j X$$

Proof of Proposition 4.1. Suppose cr satisfies (PX). That is, $cr(- \mid e_i) = C_i(- \mid E)$. Then

$$cr'(x \mid f_i) = cr(x \mid E' f_i) = cr\left(x \;\middle|\; \bigvee_{e_j, X \in \mathcal{E}_j : D_i(-) = C_j(- \mid X)} E' e_j X\right)$$

But, for each $e_j, X \in \mathcal{E}_j$ such that $D_i(-) = C_j(- \mid X)$,

$$cr(x \mid E' e_j X) = C_j(x \mid E' X) = D_i(x \mid E').$$

Now, if $c(X|A) = c(X|B)$ for mutually exclusive A, B, then $c(X|A \vee B) = c(X|A) = c(X|B)$. So

$$cr\left(x \;\middle|\; \bigvee_{e_j, X \in \mathcal{E}_j : D_i(-) = C_j(- \mid X)} E' e_j X\right) = D_i(x \mid E')$$

So $cr'(x \mid f_i) = D_i(x \mid E')$, as required.

Proof of Proposition 4.2. Suppose that cr satisfies (NX). That is, $cr(- \mid e_i) = C_i(- \mid E e_i)$. And suppose that, for each D_i, there is exactly one e_i and exactly one $X \in \mathcal{E}_i$ such that $D_i(-) = C_i(- \mid X)$. Then $f_i \equiv X e_i$. So

$$cr'(x \mid f_i) = cr(x \mid E' f_i) = cr(x \mid E' X e_i) = C_i(x \mid E' X e_i) = D_i(x \mid E' f_i)$$

as required. □

References

Peter M. Brown. Conditionalization and expected utility. *Philosophy of Science*, 43(3):415–419, 1976.

M. Burgman, M. McBride, R. Ashton, A. Speirs-Bridge, L. Flander, B. Wintle, F. Fidler, L. Rumpff, and C. Twardy. Expert status and performance. *PLoS One*, 6:e22998, 2011.

David Christensen. Epistemic Self-Respect. *Proceedings of the Aristotelian Society*, 107(3):319–336, 2007.

David Christensen. Rational reflection. *Philosophical Perspectives*, 24: 121–140, 2010.

Adam Elga. Reflection and disagreement. *Noûs*, 41:478–502, 2007.

Adam Elga. The puzzle of the unmarked clock and the new rational reflection principle. *Philosophical Studies*, 164:127–139, 2013.

Gareth Evans. *The Varieties of Reference*. Oxford University Press, 1982.

Richard Feldman. Reasonable religious disagreements. In Louise M. Antony, editor, *Philosophers without Gods: Meditations on Atheism and the Secular Life*. Oxford University Press, Oxford, 2007.

Ned Hall. Correcting the Guide to Objective Chance. *Mind*, 103:505–518, 1994.

Ned Hall. Two Mistakes About Credence and Chance. *Australasian Journal of Philosophy*, 82(1):93 – 111, 2004.

Jenann Ismael. Raid! Dissolving the Big, Bad Bug. *Noûs*, 42(2):292–307, 2008.

Jenann Ismael. In Defense of IP: A Response to Pettigrew. *Noûs*, DOI: 10.1111/nous.12057, ta.

David Jehle and Branden Fitelson. What is the 'equal weight' view? *Episteme*, 6:280–93, 2009.

A. Koriat. When are two heads better than one and why? *Science*, 336: 360–2, 2012.

Keith Lehrer and Carl Wagner. *Rational Consensus in Science and Society*, volume 24 of *Philosophical Studies Series in Philosophy*. D. Reidel, Dordrecht, 1981.

David Lewis. A Subjectivist's Guide to Objective Chance. In Richard C. Jeffrey, editor, *Studies in Inductive Logic and Probability*, volume II. University of California Press, Berkeley, 1980.

Richard Moran. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press, Princeton, 2001.

Richard Pettigrew. What Chance-Credence Norms Should Not Be. *Noûs*, DOI: 10.1111/nous.12047, ta.

Brian Skyrms. Higher order degrees of belief. In D. H. Mellor, editor, *Prospects for Pragmatism*, pages 109–137. Cambridge University Press, Cambridge, UK, 1980.

Michael Thau. Undermining and Admissibility. *Mind*, 103:491–504, 1994.

Michael G. Titelbaum. Rationality's fixed point (or: In defense of right reason). *Oxford Studies in Epistemology*, ta.

Bas C. van Fraassen. Belief and the Will. *Journal of Philosophy*, 81:235–56, 1984.

Roger White. Epistemic permissiveness. *Philosophical Perspectives*, 19: 445–459, 2005.

Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, 2000.